SPSS Guide 2009

Content

SPSS is a software package used for conducting statistical analyses, manipulating data, and generating tables and graphs that summarize data. Statistical analyses range from basic descriptive statistics, such as averages and frequencies, to advanced inferential statistics, such as correlation and regression.

There are several different SPSS windows, but the two main windows are the **Data Editor** and the **Output Viewer windows**.

- ⇨ The Data Editor is the window that is open at start-up and is used to enter and store data in a spreadsheet format.

- ⇨ The Output Viewer opens automatically when you execute an analysis or create a graph using a dialog box. The Output Viewer contains the results of all statistical analyses and graphical displays of data.

1. Basic Steps for Data Analysis

1. Get your data in SPSS. You can open a previously saved SPSS data file or enter your data directly.
2. Select a procedure. Use the menus.
3. Select the variables for the analysis. Use the dialog box for the procedure.
4. Run the procedure and look at the results. Results are displayed in the Output Viewer.

SPSS Files

1. Data Editor files are automatically saved as ****.sav** format
2. The Output Viewer files are automatically saved as *****.spo** format

In addition to files saved in SPSS-format (*.sav), you can also open and save **Excel files.**

2. Data Editor

2.1. Menus

Various pulldown menus appear at the top of the Data Editor window. Important menus are:

=>FILE used to open and save data files
=> EDIT used to copy and paste data values; used to find data in a file
=> VIEW user can change toolbars; value labels can be seen in cells instead of data values
=> DATA insert variables and cases; select, sort or weight cases; merge files
=> TRANSFORM compute new variables, recode variables, etc.
=> ANALYZE perform various statistical procedures
=> GRAPHS create bar and pie charts, etc

2.2. Features

The data editor provides a spreadsheet-like method for creating and editing SPSS data files:

1. Rows are cases. Each row represents a case or an observation. For example each individual respondent to a questionnaire is a row.
2. Columns are variables. Each column represents a variable or characteristic being measured.
3. Cells contain values. Each cell contains a single value of a variable for a case. Cells contain only data values (usually numbers).

2.3. To open an existing SPSS file

File => Open => select from the dialog box the format of the file (sav,xls..)

2.4.To create a new SPSS file

The tabs

"Data View" is the spreadsheet into which values (numbers) are entered.
"Variable View" shows the variable information (names, labels, etc.)

Variable View is used for defining new variables.

The following information is essential:

1. Name

Use SHORT name like gender, salary, var1, V1etc.

2. Type

Normally numeric is used

3. Label

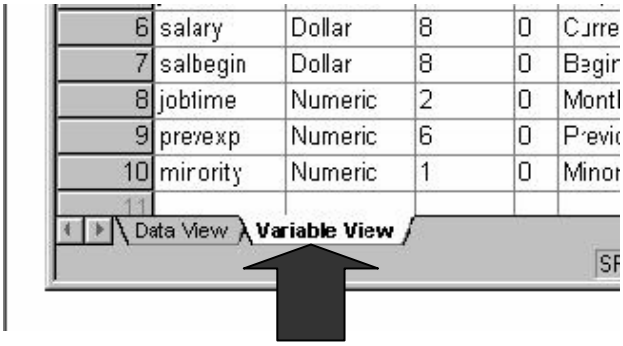
256 characters can be used to describe the variable. The labels are displayed in output results

4. Values

If the variable is coded by numbers, these codes need to have labels.
EXAMPLE: 1=male, 2=female.

5. Missing

If there are some respondents who didn't answer every question, these cells are normally empty in the Data View and no definition is needed here.



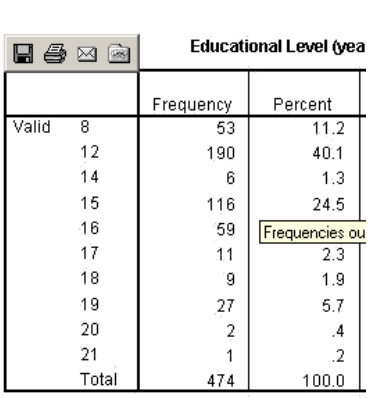
Note !
Most of the information can be **copied and pasted** from one variable to other one

Note !
The data can be edited by
1. Changing the values
2. Adding and deleting cases
4. Adding and inserting variables => ***File => Data => Insert Variable => ….***

3. Data Analysis/ Frequencies

To tabulate the number of times each value of a variable occurs. A "frequency analysis" is a way to summarize data and is referred to as a descriptive statistic. FREQUENCIES displays the results in an easy-to-read table.

Analyze => Descriptive Statistics =>Frequencies =>Select the variables...



Educational Level (yea

		Frequency	Percent
Valid	8	53	11.2
	12	190	40.1
	14	6	1.3
	15	116	24.5
	16	59	Frequencies ou
	17	11	2.3
	18	9	1.9
	19	27	5.7
	20	2	.4
	21	1	.2
	Total	474	100.0

Note !
The optionally “Charts” choice request bar chart, pie chart or histogram for chosen variables

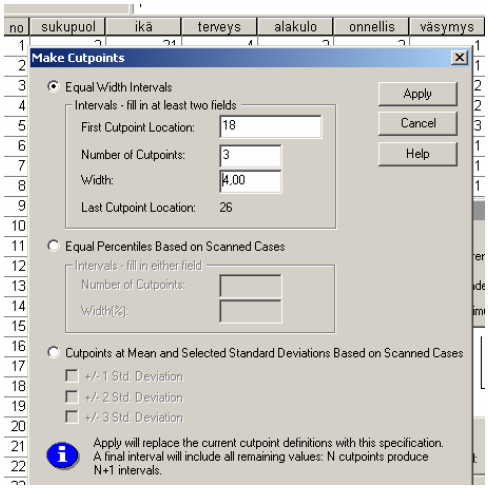
Note !
The optionally “Statistics” choice request some descriptive numbers for the chosen variables (Quartiles, Standard deviation,Range, Minimum/Maximum, Mean,Median, Mode)

4. Recoding the variable into classes

Another common type of frequency table which shows the number of data items for some variable that falls into a numerical intervals, classes.

EXAMPLE
We want to create three age categories
18-24
25-34
35-49

1. Transform => Visual Bander => …

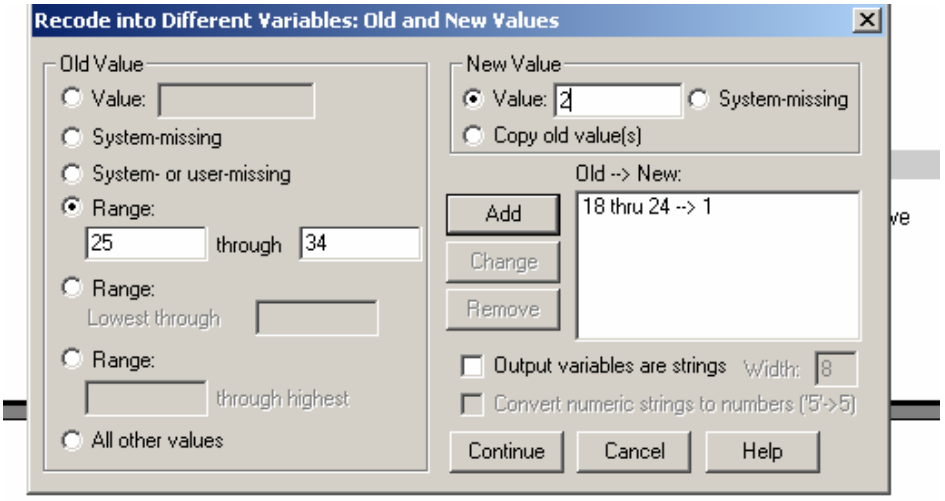


1. Chooose the variable
2. Give Banded Variable Name
2. Make Cutpoints
   => Give First cutpoint location
   => Give number of cutpoints
3. Make labels

2. Transform =>Recode =>Into different variable => .

This means that you create new variables which take the new values you assign, you save the new variables along with previously existing ones.

=> Select old variable (age) and name the Output Variable (ageclass) and the label (Age in classes)
=> **Change**
=> **Old and new values**
=> give the **Range** (25 through 34) and the class number (2) and select **Add**
=> **Continue**
=> **OK**



You have a new variable (ageclass) in your data, check the last column.

Using the variable View window you may define Value Labels once more.

5. Data Analysis/ Descriptive Numbers

Two important measures to summarize the data are:
1) Measure of location to show where the centre of the data is located
2) Measures of spread to show how spread out the data are around the centre

The following Menu choices are available:

1. Analyze => Descriptive Statistics =>Descriptive =>...

Descriptive Statistics

	N	Minimum	Maximum	Mean	Std. Deviation
salary	230	$16,200	$135,000	$35,653	$18,124

2. Analyze => Descriptive Statistics =>Explore =>Select the variables...

- ⇨ **Select the variable to explore as a dependent variable (SALARY)**
- ⇨ **Optionally you may choose a factor variable as well (GENDER)**

Descriptives

	gender		Statistic
salary	f	Mean	$26,593
		Median	$25,125
		Std. Deviation	$7,156
	m	Mean	$42,143
		Median	$33,300
		Std. Deviation	$20,671

6. Data Analysis/ Subgroup Means

The means procedure calculates **subgroup means** within categories of one or more independent variables

Analyze => Compare Means =>Means =>Select the variables...

Independent =>Category variable (GENDER)
Dependent variable => Explore variable (SALARY)

Report

Education

gender	Mean	N	Std. Deviation
f	12,78	96	2,290
m	14,34	134	2,894

7. Data Analysis/ Crosstabulation

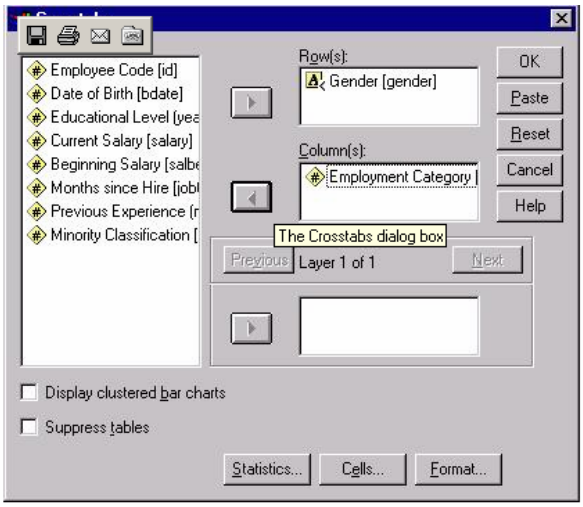
Cross tabulation table investigates if there is **any relationship** or association between two variables. Cross tabulation is a suitable method when **at least other variable is categorical** (GENDER).

Analyze => Descriptive Statistics =>Crosstabs =>Select the variables...

Rows => Dependent variable (SALARY CLASS)

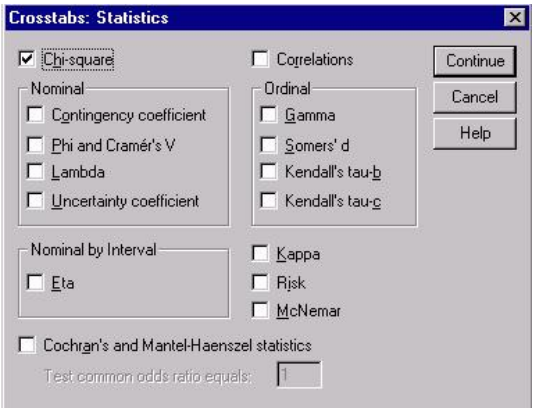
Columns => Independent variable (GENDER) Select also

Cells …..=> Percentages in columns



The Chi-square test

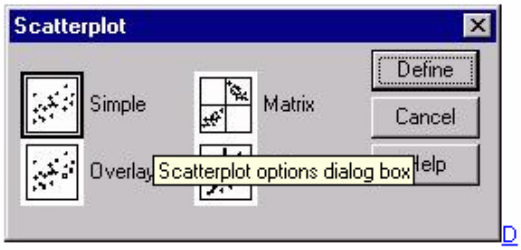
The Chi-square test of independence can be obtained through the Crosstabs dialog boxes that were used above to get a crosstabulation of the data. After opening the Crosstabs dialog box as described in the preceding section, click the **Statistics button** to get the following dialog box:



8. Data Analysis/ Scatter diagram

Scatterplots give you a tool for visualizing the relationship between two or more variables. Scatterplots are especially useful when you are examining the relationship between continuous varaibles using statistical techniques such as correlation or regression

Graphs => Scatter => Simple => Select variables



9. Data Analysis/ Correlations

When both variables are quantitative, the coefficient of correlation answers the basic question: **are the variables x and y linearly related?**

=> It means that both x and y variable are measured by numerical scale.

y = dependent variable
x = independent variable

The correlation procedure computes Pearson's correlation coefficient.

Analyze => Correlate => Bivariate => Select variables =>

10. Data Analysis/ Regression

This procedure is used to find out the equation of the best fitting line through the points on a scatter plot. Regression also allows to determine how well one variable can be used to predict another.

Correlations

		Educational Level (years)	Current Salary
Educational Level (years)	Pearson Correlation	1.000	.661**
	Sig. (2-tailed)	.	.000
	N	474	474
Current Salary	Pearson Correlation	.661**	1.000
	Sig. (2-tailed)	.000	.
	N	474	474

Analyze => Regression => Linear => Select variables =>

y = dependent variable
x = independent variable

The line best fit is found form the definitions **y = a + bx**

SPSS calculates the values of a and b.

Coefficients[a]

Model		Unstandardized Coefficients		Standardized Coefficients	t	Sig.
		B	Std. Error	Beta		
1	(Constant)	-16149.7	3255.470		-4.961	.000
	Months since Hire	161.486	34.246	.095	4.715	.000
	Previous Experience (months)	-17.303	3.528	-.106	-4.904	.000
	Beginning Salary	1.768	.059	.815	30.111	.000
	Educational Level (years)	669.914	165.596	.113	4.045	.000

a. Dependent Variable: Current Salary

Model Summary

Model	R	R Square	Adjusted R Square	Std. Error of the Estimate
1	.900[a]	.810	.809	$7,465.14

a. Predictors: (Constant), Months since Hire, Previous Experience (months), Beginning Salary, Educational Level (years)

11. Data Analysis/ Select subgroups (FILTER)

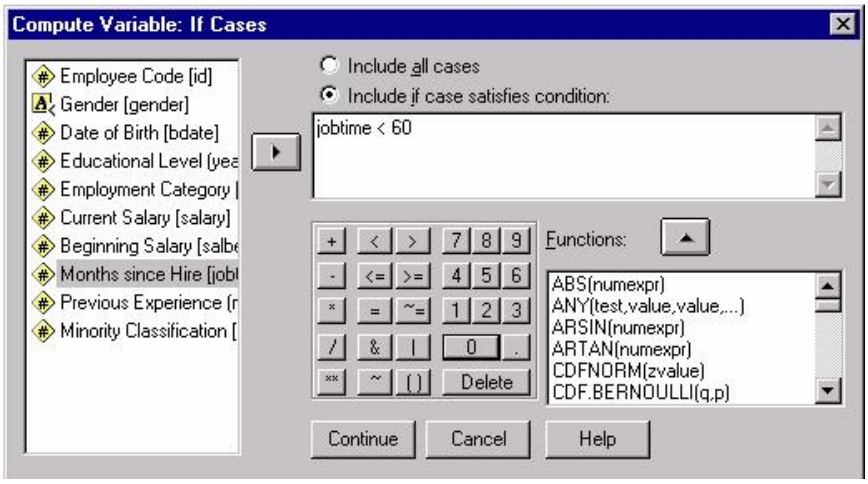
If there is a wish to focus on just one segment (subgroup) of respondents, for example, Female respondents, **the Select Cases procedure** can be used.
The procedure works as a filter for the entire data selecting only the respondents who satisfy the criteria.

Data => Select Cases =>If the condition is satisfied =>IF …

Select the variable which will be used as a criteria variable (JOBTIME)

Enter numeric expression using the different buttons under the dialog box (JOBTIME<60)

Continue
OK



When data analysis are run (frequencies, descriptive statistics, crosstabs,..) the chosen cases are active.

Note!
When you need to select all cases again, select
Data => Select Cases =>All cases

12. Charts

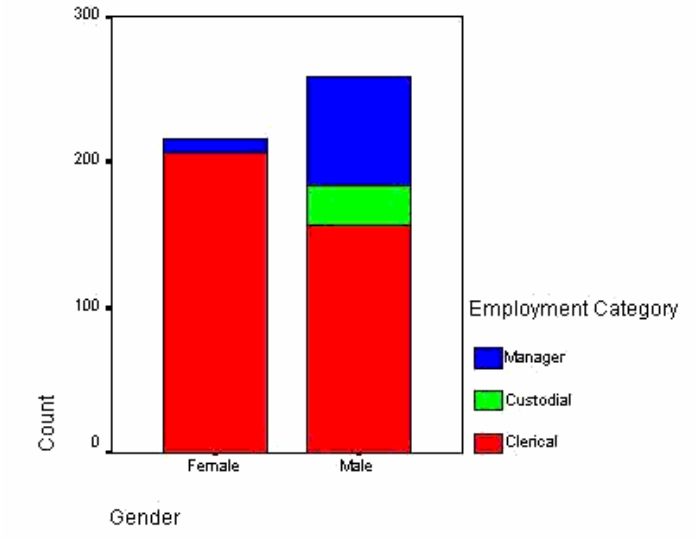
Graphs => Bar… or Pie… or Histogram…

Bar graphs

Bar Charts are a common way to graphically display the data that represent the frequency of each level of a variable.
If the bar chart is used for a **continuous variable, the *Histogram*** is used

There are three types of graphs in the dialog box:

- ⇨ The **Simple** bar graph is the most common one
- ⇨ The two other types of bar graphs, **Clustered or Stacked**, are used in situations where you want to graph frequencies for more than one variable. options in the above dialog box.



3.Output Navigator

* When the procedure is run, the results are displayed in a window called Output Navigator. In this window you can easily navigate to whichever part of the output you want to see.

** Saving output:*
*File => Save => ***.spo*

Editing Table

Double-click on the table, select from the menus =>

Format =>

1) Table Looks =>.... Templates can be used
2) Table Properties =>…Other definitions..

Editing charts

All modifications to charts are done in a chart window. You open a chart window by double clicking on a chart =>

Chart Editor menu bar is active

- ⇨ **Select the bars, pies etc. once more**
- ⇨ **Change colours**
- ⇨ **Change the scales**
- ⇨ **Change the layout**
- ⇨ **Closing the Chart Editor remain the changes**

14. Exporting SPSS Output in Other Applications

If you need to copy one output table…

1. **Select the table on the Output Navigator by clicking once on the table**
2. **Select Copy**
3. **Move to Word or Excel view**
4. **Select Paste**

You can also use Copy Objects command, then SPSS creates a picture which cannot be edited in Word or Excel.

If you need to copy one output chart …

1. Select the table on the Output Navigator by clicking once on the table
2. Select Copy Objects
3. Move to Word or Excel view
4. Select Paste

If you need to copy all output tables and charts …

1. **Select File…**
2. **Export…**
3. **Tables and Charts**
4. **All visible objects**
5. **Define the name and the place for the Word / Excel file**
6. **OK**